DOCUMENT RESUME

ED 306 270 TM 013 103

AUTHOR Green, Kathy E.
TITLE Reluctant Respondents: Differences between Early, Late and Nonresponders to a Mail Survey.
PUB DATE Mar 89
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Data Collection; Demography; *Educational Research; Elementary School Teachers; Elementary Secondary Education; Interviews; *Mail Surveys; Researchers; *Research Methodology; *Responses; Secondary School Teachers; Teacher Attitudes; Testing Problems
IDENTIFIERS Nebraska; Nonresponders; Response Rates (Questionnaires); *Teacher Surveys; Wyoming

ABSTRACT

The purpose of this study was to assess differences in demographic variables, mean attitude score, reliability of response, variability of response, and level of item completion among respondents to an initial survey mailing, to the first follow-up, to the second follow-up, and to a telephone interview. The first mailing included a cover letter, the survey instrument, and a stamped, self-addressed return envelope. Subjects were 600 inservice elementary and secondary school teachers from Wyoming and Nebraska (300 from each state). The first follow-up consisted of a reminder letter. The second follow-up consisted of a cover letter, a second copy of the survey, and a stamped, self-addressed return envelope. The first and second follow-ups were undertaken, respectively, 3 and 6 weeks following the initial mailing. Sixteen of the 600 envelopes sent out at the first mailing were undeliverable, leaving an effective sample of 584. The final response rate to the mail questionnaire was 71.2%; 25 non-respondents were interviewed by telephone. The probability of Type I error set for this study was p=0.20. Results suggest that small differences in data quality and attitude-behavior exist across waves, with reluctant respondents having fewer positive attitudes toward the topic and less favorable views of themselves as researchers and teachers. Three data tables are included. (Author/TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED306270

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
KATHY E. GREEN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Reluctant Respondents: Differences between Early, Late and Nonresponders to a Mail Survey*

Kathy E. Green
University of Denver

*Support for this project was provided by the University of Wyoming and by the Wyoming Education Association. Presented at the 1989 meeting of the American Educational Research Association, San Francisco.

TM013103

ABSTRACT

The purpose of this study was to assess differences in demographic variables, mean attitude score, reliability of response, variability of response, and level of item completion among respondents to an initial survey mailing, to the first follow-up, to the second follow-up, and to a telephone interview. The survey requested information regarding classroom teachers use of research and attitudes toward research. Subjects were 600 inservice teachers from Wyoming and Nebraska. The final response rate to the mail questionnaire was 71.2%; 25 nonrespondents were interviewed by telephone. The probability of a Type I error set for this study was p=.20. Results suggest that small differences exist across waves, with reluctant respondents having less positive attitudes toward the topic and less favorable views of themselves as researchers and teachers.

Sampling theory mandates that a random sample from a defined population be selected if population parameters are to be estimated from sample statistics. When a mail questionnaire is used as the data collection device, response rate is a primary concern since an inadequate response jeopardizes the randomness or representativeness of the sample and thus the ability to estimate population values. Consequently, perhaps the most frequently studied topic in survey research is the effect of experimental manipulations on response rate. There is less literature addressing the severity of the bias resulting from nonresponse and from delayed response--literature examining differences between early, late, and nonresponders. If respondents and nonrespondents could be systematically characterized, then bias could be predicted, at least qualitatively, and measures weighted by strata means if quantitative information were available (see Cochran, 1953: 292-304; Kish, 1965: 532-571). Similarly, if delay of response or nonresponse were found to be unrelated to parameter estimates for defined populations, less emphasis on follow-ups would be indicated when surveying those groups. While the increased numbers of responses gained from repeated follow-up would, of course, add precision to estimates by reducing standard errors, the contribution to the accuracy of central tendency measures might be minimal. The purpose of this study was to examine differences among inservice teachers who responded to an initial mailing versus those who responded to a first or second follow-up. Differences among waves in demographic variables, attitude indices, reliability of response, variability of response, and level of item completion were assessed. Differences for selected variables were also assessed for a sample of nonrespondents contacted by telephone.

Previous research has suggested two differences among waves of respondents/nonrespondents in mail surveys. Early response is associated with a higher educational level or higher social status and with more interest in the survey topic (Clausen & Ford, 1947; Dalecki, Ilvento, & Moore, 1988; Donald, 1960; Eckland, 1965; Ellis, Endo, & Armer, 1970; Franzen & Lazarsfeld, 1945; Newman, 1962; Pace, 1939; Pavalko & Lutterman, 1973; Stanton, 1939; Suchman & McCandless, 1940). However, Donald (1960) concluded that while demographic differences existed, there were no clear trends in other variables across response waves or sharp changes between respondents and nonrespondents. She argued that the telephone interviews conducted to assess nonrespondent attitudes were unjustifiable on a cost-effectiveness basis. She also concluded that while there were clear differences in interest across early waves, the minimal differences between the third and fourth waves and nonrespondents again made the use of repeated follow-ups ineffective. Other researchers found no significant differences in status of interest in the topic across waves though attitudes of nonrespondents tended to be more negative (Greenberg & Mansfield, 1957; Robinson & Agism, 1951; Toops, 1937).

Eckland (1965) found no differences in the veracity of respondents' replies across waves. His study included certified mail and telephone follow-ups with responses from 1,180 students (94% of sample). Donald (1960) and Newman (1962) found less missing data in early returns; Donald (1960) found increased variability with later

returns. Wellman, Hawk, Roggenbuck, and Buhyoff (1980), however, found no significant differences in variability or in amount of missing data across waves. It should be noted that Donald's respondent pool was 2.7 times as large as Wellman et al.'s (1,860 compared to 695) and that Donald's survey was 19 pages (198 separate responses) compared to Wellman et al.'s 12 pages (123 questions). Newman received responses to a 2-4 page questionnaire from 1100 Esquire readers. However, neither Donald nor Wellman et al. found any significant differences in mean attitude across groups and both, along with Denton, Tsai, and Chevrette (1988), Franzen and Lazarsfeld (1945), and Goudy (1976, 1978) concluded that the time and money devoted to repeated follow-ups and telephone interviews might more effectively be allocated to other activities. Wellman et al. caution, though, that if less homogeneous populations are studied, response rate would be more important. Reid (1942), however, found significant differences between the first and subsequent mailings in key variables with an overstatement of interest from early respondents compared to later returns. He dealt with a homogeneous population (school principals).

The majority of differences found between early, late, and nonrespondents involved background variables such as education and interest in the survey topic. Background variables are generally not of dominant interest in a survey, often being included for control or comparison purposes. In many situations, the existence of demographic differences across respondents may be irrelevant to the attitude or behavior being assessed. It is, however, difficult to

argue that interest in the survey topic is unrelated to attitudinal or behavioral responses.

The present study dealt with an educated, homogeneous population--inservice teachers. The sample size was smaller than those of the Donald (1960) or Wellman et al. (1980) studies and the length of the survey was substantially shorter (4 pages compared to 12-19 pages). Differences in reliability of responses across waves were examined as well as differences in variables identified by previous researchers. Null hypotheses were that no significant differences would be found across waves in demographic variables, mean attitude score, reliability of response, variability of response, or level of item completion. Post hoc comparisons were performed contrasting (1) respondents and nonrespondents and (2) Wave 1 respondents vs. respondents to follow-ups (Waves 2 and 3). Results of the former comparison addressed the utility of interviewing nonrespondents; results of the latter comparison addressed the utility of follow-up mailings.

METHOD

Sample

Subjects were inservice teachers in the States of Wyoming and Nebraska. Six hundred names were selected from the State Department of Education lists of teachers in the spring of 1987, 300 from each state. Nebraska teachers were randomly selected. Wyoming teachers were stratified by level taught (elementary, middle school, secondary) and then randomly selected (100 from each group).

A sample of nonrespondents was selected to interview by telephone the following fall. These persons were asked 12 of the key questions from the survey. To obtain a sample of 25 nonrespondents, 43 persons were telephoned. Of these, one refused to be interviewed and 17 were no longer teaching at that school. The remaining twenty-five nonrespondents were interviewed. Although the sample of nonrespondents was initially random, inability to contact all persons in the sample and inability to obtain telephone numbers for females resulted in a sample of convenience for this group.

Instrument

The survey form had 51 items on two pages, double-sided. Of these items, 46 were fixed response items and 5 were open-ended. Length of time for survey completion was estimated at approximately 15 minutes. The survey addressed teachers' opinions regarding the application of research methods and findings to classroom teaching.

The following measures were developed for respondents. Attitude indices were calculated by averaging responses across items to create four separate measures, three assessing practices in use of research and one attitude toward research measure. The three measures of use of research were review of research literature, conduct of research, and presentation of research at a professional meeting.

Procedure and Analysis

The first mailing (Wave 1) consisted of a cover letter, the survey, and a stamped, self-addressed return envelope. The first follow-up (Wave 2) was a letter of reminder; the second follow-up (Wave 3) was a cover letter with a second copy of the survey and a

stamped, self-addressed return envelope. Surveys were mailed to teachers in September, 1987. The first follow-up was sent three weeks after the initial mailing with the second follow-up sent six weeks later.

Sixteen of the 600 envelopes mailed were non-deliverable leaving an effective sample of 584. The overall response rate was 48.6% (n=286) to the first mailing, 6.5% (n=38) to the first follow-up, and 15.8% (n=92) to the second follow-up for a final response rate of 71.2% (n=416). There were two refusals.

Reliability of response was calculated using Cronbach's alpha. Variability of response was calculated for each scale separately. The level of missing data was calculated separately for open-ended and for closed response items by averaging missing responses across persons. The proportion of persons volunteering examples of their use of research was also tabulated. Differences across waves were assessed using analyses of variance, Hakstian and Whalen's (1976) test of the homogeneity of internal consistency coefficients, Cochran's test for homogeneity of variances, and tests of proportions.

Differences between respondents and nonrespondents were assessed for selected items using analysis of variance and chi-square statistics.

Since interest in this study was actually in <u>failing</u> to reject the null hypotheses, a liberal alpha level (p=.20) was used to determine significance of differences. This was done to provide greater power for detecting differences among means. Use of a

stingent alpha level would attribute weak differences across waves to random error. So, as in any situation when one is seeking to accept the null hypothesis of equality of parameters, even weak differences must be considered as an indication of the potential falsity of the null model. Thereby, a conservative approach was taken with regard to statements about the equality of parameters across waves and the resulting utility of interviews and follow-up mailings.

RESULTS

Differences among the four groups (three waves of respondents and the nonrespondents interviewed by telephone) were assessed. Significant differences were found between respondents and nonrespondents in self-rating of skill as a teacher, with nonrespondents (Mean = 4.76) rating themselves significantly lower than Wave 1 and Wave 2 respondents (Means of 5.16 and 5.35)(F=5.23, p<.01). This difference may be due in part to use of a phone interview rather than a mail questionnaire for the nonrespondent group. Teachers may have rated themselves lower when speaking to a person rather than when completing a survey. Self-rating of skill as a research reader also differed across waves (F=1.71; p<.17). Early respondents rated themselves more highly (Mean = 3.43) than nonrespondents (Mean = 2.92). There were significant differences in years of teaching (F=1.98, p<.16) with nonrespondents having fewer years of teaching. There were no significant differences in grade level taught, numbers of research courses taken, or conferences attended. No differences were found in frequency of reviewing the

literature, collecting data, or writing research reports. There was a significant difference in the proportion of males and females in the nonrespondent group, with males comprising 80% of that group and 43% of the respondent group. This point was addressed earlier: it was not possible to get telephone numbers for many women. When Waves 1, 2, and 3 were combined and contrasted to nonrespondents, one additional difference was found ($F=1.74$, $p<.19$). Nonrespondents rated themselves lower as research producers than did respondents. It seems, then, that while the reported behavior of nonrespondents did not differ from the reported behavior of respondents, nonrespondents viewed themselves as less proficient at research activities and less skilled as teachers.

In the following analyses, only Waves 1, 2, and 3 were included since the needed information was not available for the nonrespondent group. Significant relationships were found between response wave and age group ($\chi^2=16.44$, $p<.04$) and between response wave and sex ($\chi^2=6.21$, $p<.05$). Older persons tended to respond at a higher rate to the initial mailing than younger persons and females responded at a higher rate to the initial mailing than did males. (The latter result is in contrast to that found by Green and Stager, 1986, who found males to respond to the initial mailing at a higher rate than females.) No significant relationships were found for grade level taught, subject taught, or years in teaching. A higher proportion of Wave 1 respondents held graduate degrees (42.7%) than Wave 2 (36.8%) or Wave 3 (37.0%) respondents but differences were not significant.

The null hypothesis of no demographic differences across waves was, therefore, rejected for age and sex.

Differences in attitude and use of research indices across the three waves were not significant. However, when respondents to Waves 2 and 3 were pooled and contrasted with respondents to the initial mailing, significant differences were found for review of the literature, conduct of research, and attitudes toward use of research (Table 1). In all cases, attitudes were more positive for respondents to the initial mailing than for later respondents. Significant differences were also found in a self-rating item of skill as a teacher (F=1.69, p<.19) with respondents to the initial mailing rating themselves more favorably.

Internal consistency reliability estimates were calculated for each measure for each of the three waves and are presented in Table 2. Differences in reliability estimates across waves were significant for the attitudes toward research measure (M=8.32, p<.02) but not for the other three measures. No differences were found to be significant when Waves 2 and 3 were combined and contrasted to Wave 1 responses.

Scale variability was calculated for Waves 1, 2, and 3. Significant differences in variability were found for review of the research literature (Cochran's C=.41, p<.04) with the second wave being the most variable and Wave 3 the least variable. Significant differences in variability were also found for attitude toward research (Cochran's C=.44, p<.01), with the last wave being the most variable. The null hypothesis of no differences across waves in variability was rejected for two of the four measures (Table 3).

However, when Waves 2 and 3 were combined and contrasted to Wave 1, there were no significant differences in variances.

The mean number of missing responses to closed and open-ended items was tabulated and compared for Waves 1, 2, and 3. While there was a lower mean number of missing responses for closed and open-ended questions for Wave 1 than for Waves 2 and 3, differences were not significant. Respondents were asked at the end of the survey to provide examples of their use of research. A significantly higher proportion of respondents to the first (.47) and second (.45) waves provided comments than respondents to the third wave (.35)(z=2.93, p<.05).

DISCUSSION

Results of this study are consistent with the results of previous research in several ways. Since the population for this study was a homogeneous, well-educated group, no differences in level of education were found. However, differences were found across waves for other demographic variables (age and sex). The result of demographic differences is consistent with those of previous studies.

The result of no significant differences in attitude/behavior measures at p<.05 is consistent with results of some previous studies. However, differences were found at a more liberal alpha level (p<.20). These differences suggest that respondents to the initial mailing had more favorable attitudes toward the topic and also had more positive attitudes about themselves. This result is also consistent with results of some previous research studies. It

should be noted that studies finding no significant differences across waves generally used $p<.01$ or $p<.05$ as the probability of a Type I error which may in part account for the failure of those studies to find differences. Differences, at least in this study, exist but were weak. And, of course, some may be attributable to Type I errors. But, if a conversative view is taken, even weak differences cannot be ignored.

Differences in internal consistency reliability across waves have not been examined previously to the author's knowledge. Significant differences were found for reliability coefficients and also for scale variability. Since internal consistency reliability is in part a function of scale variability, these results are not independent. Where significant differences in reliability were found, differences in scale variability were found such that higher reliability was paired with higher variability. But, the pattern of differences in variability differs from that found by Donald (1960). In her study, variability increased with later returns. In the present study, variability increased with later returns for one measure (the most internally stable measure) but not for other measures. Also in contrast to Donald's results, the incidence of missing data was not significantly different across waves except for response to a solicitation for examples of the use of research. Donald found an increasing proportion of missing data across waves. One explanation of the difference between the present and Donald's results could lie in the decreased power to find differences of the present study (due to the smaller sample size employed). The incidence of missing data

did increase across waves though the differences were not significant.

In summary, the results of this study suggest that there are differences in both data quality and attitude-behavior across waves of respondents. Further, significant differences were found between respondents and nonrespondents. Only one difference was significant at $p<.01$ (the variance in attitudes toward research across waves) but at $p<.20$, delay of response seemed to be associated with less interest in the topic and lower self-perceptions of skills. The results of this study, then, do not suggest that major differences are to be found across waves of respondents but that minor differences exist. To accurately estimate a population mean, then, it seems necessary to assess the attitudes/behaviors of reluctant respondents. In agreement with Dalecki et al. (1988), results of this study suggest that the number of mailings cannot be reduced without adversely affecting accuracy of parameter estimates and generalizability. An added caution is in order. In agreement with Wellman et al. (1980), it may be even more important to follow-up nonrespondents when dealing with a less homogeneous group.

REFERENCES

Clausen, J.A., & Ford, R.N. (1947). Controlling bias in mail questionnaires. Journal of the American Statistical Association, 42, 497-511.

Cochran, W.G. (1953). Sampling techniques. New York: John Wiley.

Dalecki, M.G., Ilvento, T.W., & Moore, D.E. (1988). The effects of multi-wave mailings on the external validity of mail surveys. Journal of the Community Development Society, 19, 51-70.

Denton, J.J., Tsai, C.Y., & Chevrette, P. (1988). Effects on survey responses of subjects, incentives, and multiple mailings. Journal of Experimental Education, ,77-82.

Donald, M.N. (1960). Implications of nonresponse for the interpretation of mail questionnaire data. pUBLIC oPINION qUARTERLY, 24, 99-114.

Eckland, B.K. (1965). Effects of prodding to increase mail-back returns. Journal of Applied Psychology, 49, 165-169.

Ellis, R.A., Endo, C.M., & Armer, J.M. (1970). The use of potential nonrespondents for studying nonresponse bias. Pacific Sociological Review, 13, 103-109.

Franzen, R., & Lazarsfeld, P.F. (1945). Mail questionnaire as a research problem. Journal of Psychology, 20, 294-.

Goudy, W.J. (1976). Nonresponse effects on relationship between variables. Public Opinion Quarterly, 40, 360-369.

--- (1978). Interim response to a mailed questionnaire impacts on variable relationships. The Sociological Quarterly, 19, 253-268.

Greenberg, A., & Mansfield, M. (1957). On the reliability of mailed questionnaires in product tests. Journal of Marketing, 21, 342-345.

Hakstian, A.R., & Whalen, T.E. (1976). A K-sample significance test for independent alpha coefficients. Psychometrika, 41, 219-231.

Kish, L. (1965). Survey sampling. New York: John Wiley & Sons.

Newman, S.W. (1962). Differences between early and late respondents to a mailed survey. Journal of Advertising Research, 2, 37-39.

Pace, C.R. (1939). Factors influencing questionaire returns from former university students. Journal of Applied Psychology, 23, 388-.

Pavalko, R.M., & Lutterman, K.G. (1973). Characteristics of willing and reluctant respondents. Pacific Sociological Review, 16, 463-476.

Robinson, R.A., & Agism, P. (1951). Making mail surveys more reliable. Journal of Marketing, 15, 415-424.

Stanton, F. (1939). Note on the validity of mail questionnaire returns. Journal of Applied Psychology, 23, 95-104.

Suchman, E.A., & McCandless, B. (1940). Who answers questionnaires? Journal of Applied Psychology, 24, 758-767.

Wellman, J.D., Hawk, E.G., Roggenbuck, J.W., & Buhyoff, G.J. (1980). Mailed questionnaire surveys and the reluctant respondent: An empirical examination of differences between early and late respondents.

Table 1

Differences in Use of and Attitude toward Research for Respondents to the Initial vs. Follow-up Mailings

Measure	Initial Mailing Mean	SD	n	Subsequent Mailings Mean	SD	n	F	p
Review of research literature	3.47	1.18	284	3.25	1.18	129	3.18	.08
Conduct of research	1.85	.96	285	1.68	.88	129	3.03	.08
Attitude toward use of research	3.81	.50	282	3.73	.52	126	2.22	.14

Table 2

Reliability Estimates by Response Wave

Measure	n items	Wave 1	2	3	M	p
Review of research literature	5	.81	.86	.79	.19	.90
Conduct of research	3	.72	.76	.77	1.23	.70
Presentation of research at a professional meeting	3	.75	.74	.76	1.98	.60
Attitude toward use of research	23	.75	.67	.84	8.32	.05
n cases		284	37	92		

Table 3

Scale Standard Deviations by Response Wave

Measure	Wave 1	Wave 2	Wave 3	Cochran's C	p
Review of research literature	1.18	1.35	1.11	.408	.04
Conduct of research	.96	1.06	1.04	.368	.43
Presentation of research at a professional meeting	.78	.59	.81	.404	.06
Attitude toward use of research	.50	.38	.56	.437	.01